Usability Testing as Evaluation: Development of a Tool

Charlene Weir, Ph.D., Michael J. Lincoln, M.D, and JoAnn Green

INTRODUCTION

Software is usually designed to replace a manual method or enhance a current practice. Usability testing is a process by which user acceptability and software function can be measured. Such testing borrows from human factors, engineering, and cognitive psychology. During a usability test, the user performs a specified task with a version of the software. The user receives minimal instruction and is told to "think aloud" during the process. Testers observe and record the user's actions and speech, but do not interact with the user. They note the user's choice of methods to perform the tasks, any problems which are encountered, and how the user works around the problems. This poster presents a taxonomy of usability testing dimensions developed by the Usability Testing Team at the Salt Lake City Veterans Affair's Information Resources Management Field Office. The dimensions we report provide a method of measuring and categorizing specific aspects of usability, and can be used to compare the same software over time and compare across applications.

METHODOLOGY

The VA Computerized Patient Record System v. 1.0 (CPRS) application was the focus of the usability testing for this study. The CPRS was programmed in Borland's Delphi and was designed to provide a PC-based, graphical user interface for order entry, results reporting, and clinician communication. The testing included 22 test subjects (physicians). Each physician was given two clinical scenarios (based on actual patient charts) asking them to admit, write orders, progress notes, enter problems, check results and discharge. For each session, all problems were documented and later reviewed in order to list all unique problems.

RESULTS

The dimensions listed in Table 1 were derived from the observed usability problems categorized according to the available literature in usability and human cognition. After discussion and adjustment, each observed problem was coded independently by three coders. Any particular problem could be given more than one code (hence the table sums to more than 100%). The resultant level of agreement between coders was 85%.

TABLE 1. Usability Dimensions and % occurrence across sample problems.

DIMENSION	%
Task Content Representation (e.g. list of drugs is complete)	20%
Task Sequence Representation (e.g., it is easy to order drugs and then write note)	8%
Lack of Feedback (i.e., system echoes input, identification of position errors)	24%
Cueing (i.e., clarity of icons, precision of menus, and helpful help functions)	32%
Visibility (e.g., screen too crowded; takes a long time to find an item)	8%
Failure to allow appropriate output or displays (e.g., correlating lab data & meds)	4%
Failure to meet protocols (either VA internal protocols or industry)	52%
Failure to support necessary multiple activities (e.g., look up drug while writing note)	0%

SUMMARY

Simple usability assessments (e.g., time for task, attitude surveys, simple counts of user problems) do not often provide specific, useful insights. We propose that usability testing, categorized by cognitively-based dimensions, can offer better information. For example, testing may show that users reproducibly fail to visualize the proper menu choice or button to carry out a task. This may occur because the interface tends to direct the user's attention to a non-relevant menu or interface widget. The software may also provide inadequate feedback on task completion: one might observe a user entering a medication order several times in succession, while indicating aloud that they weren't sure it was "taken off" by the computer. We have found that, using this type of information, developers are able to make both specific improvements (e.g., fix a confusingly worded dialog) and general classes of improvements (e.g., simplify the interface, better utilize users' existing mental models for the task, provide interface features which direct the user's attention to relevant tasks at particular times). The usability dimensions further provide a means to compare different versions of the same application as well as different applications developed for the same task.

0195—4210/96/$5.00